Clan CD of cysteine peptidases as an example of evolutionary divergences in related protein families across plant clades

Ines Cambra, Francisco J. Garcia, Manuel Martinez

Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Campus Montegancedo, 28223-Pozuelo de Alarcón (Madrid), Spain

A B S T R A C T

Comparative genomic analyses are powerful tools that can be used to analyze the presence, conservation, and evolution of protein families and to elucidate issues concerning their function. To deal with these questions, we have chosen the clan CD of cysteine peptidases, which is formed by different protein families that play key roles in plants. An evolutionary comparative analysis of clan CD cysteine peptidases in representative species of different taxonomic groups that appeared during the evolution of the Viridiplantae was performed. The results obtained indicates: i) C13 GPI:protein transamidases, C14 metacaspases I, and C50 separases are present in all taxonomic groups; ii) C13 legumains and C14 metacaspases II are absent in some basal algae groups; iii) C11 clostripains have only been found in the two Chlorophyceae species; iv) C25 gingipains and C80 RTX toxins have not been found in plants. Moreover, gene duplication events could have been associated in some families to the increasing complexities acquired in land plants. These findings have demonstrated that comparative genomics is useful to provide valuable insights on the differential evolution of the related peptidase families belonging to clan CD in plant clades. The low number of protein members suggests a restricted physiological role for these peptidase families, mainly in algae species.

Keywords:
Algae
Comparative genomics
Cysteine peptidase
Molecular evolution
Protein families
Land plants

1. Introduction

Extensive genome sequencing has released a large amount of data available to perform comparative genomic analyses in different plant clades. As a consequence, valuable insights into the conservation and evolution of a protein family can be obtained, which could aid in elucidating issues concerning the function of these proteins. As an example of how evolution in different, but related, protein families can be inferred using this tool, we selected the clan CD of cysteine peptidases, which are enzymes that hydrolyse peptide bonds using a catalytic cysteine. The MEROPS database (Rawlings et al., 2008) contains all the modern-day peptidases grouped in clans. Clans represent one or more families that show evidence of their evolutionary relationship by their similar tertiary structures, or when structures are not available, by the order of catalytic-site residues in the polypeptide chain and often by common sequence motifs around the catalytic residues. At present, there are 72 families of cysteine peptidases: 43 families are included in 9 clans exclusively formed by cysteine peptidases (CA, CD, CE, CF, CH, CL, CM, CN, CO), 13 families are included in 3 clans that comprise peptidases with different catalytic mechanisms (PA, PB, PC), and 16 families are not enclosed in any determined clan. Most reported plant cysteine peptidases are located in clan CA. However, in the last years, an increased number of cysteine peptidases from clan CD have been characterized. Clan CD is composed by 6 protein families: C11 (clostripains), C13 (asparaginyl endopeptidases or legumains and GPI:protein transamidases), C14 (formed by subfamilies C14A, caspases; and C14B, paracaspases and metacaspases I and II), C25 (gingipains), C50 (separases), and C80 (RTX toxins). These protein families share a protein fold or similar sequence motifs. All families contain a His, Cys catalytic dyad. The catalytic His occurs in a His-Gly motif and is preceded by a block of hydrophobic residues; the catalytic Cys is preceded by a second block of hydrophobic residues (Chen et al., 1998). Tertiary structures have been determined for members of families C14 (Walker et al., 1994), C25 (Eichinger et al., 1999), and C80 (Lupardus et al., 2008). These show alpha/beta proteins with a fold that consists an alpha/beta/alpha sandwich. Other families are included in the clan because of the conservation of motifs around the catalytic residues (Chen et al., 1998). Specificity is strongly directed to the P1 residue of the substrate, which is normally Asn or Asp in family C13, Asp or Arg in family C14, and Arg (or sometimes Lys) in C11, C25, and C50.

To date, only members of the C13, C14, and C50 families have been described in plants. C13 family is formed by legumains or VPE (vacuolar processing enzymes) and GPI:protein transamidases. There

are abundant evidences indicating that legumains perform a protein-processing function that causes a limited proteolysis of precursor proteins (Hara-Nishimura et al., 1991). Legumain from plant seeds is thought to be responsible for the post-translational processing of seed proteins prior to storage (Shimada et al., 2003). During germination, legumains contribute to the activation of cysteine peptidases to degrade storage proteins (Okamoto and Minamikawa, 1999; Kato et al., 2003). A role in defense against pathogens executing programmed cell death due to the caspase activity observed for several legumains has also been proposed (Hatsugai et al., 2004; Rojo et al., 2004). Recently, a comprehensive review of legumains in different plant clades has been published compared to their proteinaceous inhibitors (Martinez and Diaz, 2008). GPI:protein transamidases are the catalytic subunits of a protein complex described in yeast, mammals, and parasitic protozoa involved in the attachment of a glycosylphosphatidylinositol (GPI) to proteins destined to be anchored to the plasma membrane. These proteins have a N-terminal signal sequence directing them to the endoplasmic reticulum and a C-terminal signal that directs cleavage of a propeptide and replacement by a GPI anchor by an acyl transferase reaction that forms a peptide linkage between the terminal amine of the ethanolamine phosphate group of the GPI anchor and the C-terminal carbonyl group of the protein (reviewed in Zacks and Garg, 2006). In plants, orthologs of GPI:protein transamidases have been described and their putative target proteins in silico predicted (Eisenhaber et al., 2003a,b).

The family C14 includes caspases, paracaspases, and metacaspases I and II. In plants, only metacaspases have been described. Metacaspase I sequences have been reported in fungi, protozoa, and plants (Uren et al., 2000; Bidle and Falkowski, 2004; Vercammen et al., 2007). Metacaspase II sequences have been restricted to plants, with the exception of a metacaspase II in the protozoan Monosiga probably acquired by horizontal transfer from a green algae (Nedelcu et al., 2008). The assignment of metacaspases and caspases to the same family is controversial, since the P1 preference of metacaspases is basic (Vercammen et al., 2004; Gonzalez et al., 2007), whereas that of caspases is acidic, and previous phylogenetic analyses of clan CD peptidases have shown that caspases and metacaspases constitute separate groups (Aravind and Koonin, 2002). This suggests different functions for both kinds of proteins. In fact, although metacaspases have been involved in responses to stress (Bidle and Falkowski, 2004; Belenghi et al., 2007), the role of metacaspases in cell death, which is the main function of most animal caspases (Chowdhury et al., 2008), remains enigmatic. Caspase-like activity reported in plant and fungi cell death could be exerted by other proteases exhibiting caspase-like activity (Hatsugai et al., 2006).

The family of separases (C50) was originally discovered by genetic analysis of mitosis in fungi (Baum et al., 1988). These genes encode large proteins with conserved sequences near the C-termini that were recognised as homologous to peptidases in clan CD (Uhlmann et al., 2000). Separase has been shown to be required for the separation of sister chromatids during mitosis and meiosis in a range of organisms from yeasts to *Arabidopsis* and vertebrates by cleavage of the cohesin subunit Scc1 (Queralt and Uhlmann, 2005; Liu and Makaroff, 2006).

As previously stated, comparative genomic analyses could provide valuable insights into the conservation and evolution of these protein families. Thus, we have performed extensive searches and phylogenetic analyses of the different clan CD peptidases in representative species of different taxonomic groups belonging to Viridiplantae. The results indicate that whereas C13 GPI:protein transamidases, C14 metacaspases I, and C50 separases are present in all taxonomic groups, C13 legumains and C14 metacaspases II are absent in some basal groups. Moreover, for first time, C11 clostripains have been detected in some algae species. In some families, several gene duplication events could have been associated to the increasing structural and functional complexities acquired in land plants.

2. Material and methods

2.1. Databases searches

BlastP and TBlastN searches for clan CD cysteine peptidases were performed in publicly available genome databases and in The J. Craig Venter Institute (JCVI) plant transcript assemblies (TA) database (http://plantta.jcvi.org/index.shtml) which was built from expressed transcripts collected from dbEST (ESTs) at the NCBI GenBank nucleotide database. Sequences for *Oryza sativa* ssp. japonica (rice annotation release 5) and *Ricinus communis* (release 1) were obtained at JCVI (http://www.jcvi.org). Sequences for *Arabidopsis thaliana* were identified by searching The Arabidopsis Information Resource (TAIR) database (TAIR7 genome release; http://www.arabidopsis.org). Searches for algae, moss, spikemoss, poplar, and sorghum sequences were carried out at the DOE Joint Genome Institute (JGI; http://www.jgi.doe.gov), using the current releases: *Chlamydomonas reinhardtii* v3.0; *Volvox carteri* f. nagariensis v1.0; *Chlorella vulgaris* strain C-169 v1.0; *Chlorella* sp. strain NC64A v1.0; *Ostreococcus lucimarinus* v2.0; *Ostreococcus* sp. strain RCC809 v1.0; *Ostreococcus tauri* v2.0; *Micromonas pusilla* strain CCMP1545 v2.0; *Micromonas* sp. strain RCC209 v2.0; *Physcomitrella patens* ssp. patens v1.1; *Selaginella moellendorffii* v1.0; *Populus trichocarpa* v1.1; *Sorghum bicolor* v1.0. Blast searches were made in a recurrent way. First, if available, a complete amino acid plant sequence from data banks corresponding to a protein of the family was used. If not, we used a protein belonging to the family of any other organism. Then, the protein sequences of each plant species were used to search in the species. Finally, after an alignment of the proteins found in plants, the conserved region surrounding the catalytic sites from the species most related was used to a final search in each plant species.

Information about gene models for all these proteins is compiled in Supplementary data 1. Additionally, Psi-Blast searches in general protein databases were made. As a result, any additional group of cysteine peptidases putatively belonging to clan CD was not found.

2.2. Protein alignments and phylogenetic trees

Alignments of the amino acid sequences were performed using the default parameters of MUSCLE version 3.6 (Edgar, 2004). Depicted alignments were obtained by the multiple alignment editor Jalview version 2.4 (Waterhouse et al., 2009). Alignments ambiguities and gaps were excluded from phylogenetic analysis using GBLOCKS version 0.91b (Castresana, 2000). Phylogenetic and molecular evolutionary analyses were conducted using the programs PhyML (Guindon and Gascuel, 2003) and MEGA version 4.0 (http://www.megasoftware.net; Guindon and Gascuel, 2003; Tamura et al., 2007). The program PROTTEST (2.2) was employed for selecting the model of protein evolution that fits better to each alignment according to the corrected Akaike Information Criterion (Abascal et al., 2005). The parameters of the selected models were employed to reconstruct the displayed clan CD cysteine peptidases trees by means of a maximum likelihood PhyML method using a BIONJ starting tree. The approximate likelihood ratio test (aLRT) based on a Shimodaira–Hasegawa-like procedure was used as statistical test for nonparametric branch support (Anisimova and Gascuel, 2006). Trees were rooted using as outgroup protein sequences of the same family belonging to non-plant species. All families were also analyzed with the maximum parsimony and the neighbour-joining algorithms and with different gap penalties. No significant differences in the tree topologies were detected.

3. Results

3.1. Number of clan CD cysteine peptidases in completely sequenced plants

Nine Chlorophyta algae (five Prasinophyceae, *M. pusilla* CCMP1545, *Micromonas* sp. RCC209, *O. tauri*, *O. lucimarinus*, and *Ostreococcus* sp.

RCC809; two Trebouxiophyceae, *C. vulgaris* C-169, *Chlorella* sp. NC64A; and two Chlorophyceae, *C. reinhardtii* and *V. carteri*), one moss (*P. patens*), one spikemoss (*S. moellendorffii*), and five angiosperms (three dicots, *A. thaliana*, *P. trichocarpa*, and *R. communis*; and two monocots, *O. sativa* and *S. bicolor*), which have been completely sequenced and drafts of these sequences are available on the Web, were selected to establish the number of clan CD cysteine peptidases in each species. Additionally, public plant EST collections were searched. Sequence similarity surrounding the putative location of the conserved catalytic histidine and cysteine was used to include each protein in a specific CD cysteine peptidase family. C14 caspases, C25 gingipains, and C80 RTX toxins were not detected in any plant species, neither in the selected genomes nor in public EST collections, with the exception of a truncated protein homologous to bacterial paracaspase-like proteins found in *O. tauri*. The results obtained from genome extensive searches for the rest of families of CD cysteine peptidases are summarized in Table 1. All the algae species present C13 GPI:protein transamidases and C14 metacaspases I. C14 type II metacaspases were not found in the Prasinophyceae and Trebouxiophyceae species. C13 legumains were lacking in Prasinophyceae and *Micromonas* species were devoid of C50 separases. In the land plants, the number of C13 legumains and C14 metacaspases I and II was higher than in algae, whereas the number of C13 GPI:protein transamidases and C50 separases did not increase. Additionally, C11 clostripains were detected in the two species of Chlorophyceae algae.

3.2. Evolution of C13 cysteine peptidases in plants

The C13 family of cysteine peptidases includes legumain-like peptidases and GPI:protein transamidases. To obtain further insights on how this family has evolved from algae to angiosperms, the C13 complete proteins were aligned by MUSCLE (see Supplementary data 2), and phylogenetic trees were constructed by the maximum likelihood PhyML method. The truncated legumain model from *V. carteri* (VcLeg-1) was excluded from this analysis since it lacks the region surrounding the catalytic Cys residue. The rest of legumains and GPI:protein transamidases conserved the catalytic histidine and cysteine residues (Figs. 1a and b). However, the amino acid residues surrounding the putative location of the His and Cys catalytic residues were specific for each protein type. Legumains have typical DHG and ACE motifs, whereas GPI:protein transamidases have GHG and TCQ motifs. The corresponding phylograms are shown in Fig. 2. The C13 legumains from Chlorophyceae and Trebouxiophyceae were the most divergent and can be considered ancestors to the other sequences. The rest of the legumain sequences are grouped in three different clades supported by approximate likelihood ratio test values (aLRT) higher than 80%. The first clade includes proteins from the moss, the spikemoss, and all the angiosperm species. The second clade is exclusively formed by angiosperm species. The third clade contains one sequence from the species *P. trichocarpa* (PtLeg-5) and *A. thaliana* (AtLeg4). Orthologous genes (different genes that originated in speciation events) and paralogous genes (different genes that originated in a duplication event) could be detected for proteins located in the different groups. For example, a duplication event in the ancestor of angiosperm species originated the protein from which the sequences in the second clade were derived. Then, the sequences in the first and second clades are paralogs. In each clade, orthologous relationships can be deduced to several proteins. In the first and second clades, species-specific duplications in most species make it difficult to assign actual orthologs, but orthology can be assumed for the cereal proteins SbLeg-1 and OsLeg-1, SbLeg-2 and OsLeg-2, or SbLeg-3 and OsLeg-3. The phylogram for GPI:protein transamidases resembles the evolutionary relationships among plant species, from algae to angiosperms, which denotes the ancestral characteristic of this kind of proteins and indicates that the genes are orthologs. The Prasinophyceae sequences were the most divergent and grouped with the Chloropyceae and Trebouxiophyceae proteins. The rest of the sequences are located in a second clade, in which the angiosperm GPI:protein transamidases are closely related.

Table 1
Number of different clan CD cysteine peptidases from algae to land plants.

Clade	Organism	C11 (Clost)	C13 (Leg)	C13 (GPI)	C14 (MCI)	C14 (MCII)	C50 (Sep)
Prasinophyceae	*M. pusilla* CCMP1545	0	0	1	1	0	0
	Micromonas sp. RCC299	0	0	1	1	0	0
	Ostreococcus sp. RCC809	0	0	1	1	0	1
	O. tauri	0	0	1	1	0	1
	O. lucimarinus	0	0	1	1	0	1
Trebouxiophyceae	*Chlorella* sp. NC64A	0	1	1	1	0	1
	C. vulgaris C-169	0	1	0	3	0	1
Chlorophyceae	*C. reinhardtii*	1	1	1	1	1	1
	V. carteri	1	1	1	1	1	1
Bryophyta	*P. patens*	0	4	1	2	5	1
Lycopsida	*S. moellendorffii*	0	2	1	2	1	1
Monocotyledons	*O. sativa*	0	5	1	5	4	1
	S. bicolor	0	4	1	5	7	1
Eudicots	*A. thaliana*	0	4	1	3	6	1
	P. trichocarpa	0	5	1	8	4	1
	R. communis	0	2	1	7	2	1

Clost, clostripain; Leg, legumain; GPI, GPI:protein transamidase; MCI, metacaspase I; MCII, metacaspase II; Sep, separase.

3.3. Evolution of C14 cysteine peptidases in plants

Metacaspases I and II can be differentiated by their different protein structure. Metacaspases I have an N-terminal domain before the p20 subunit that is not present in metacaspases II. Likewise, metacaspases II have a spacer region between the subunits p20 and p10 longer than metacaspases I (Fig. 3a). The catalytic His and Cys residues are in the p20 subunit separated by 50–55 amino acid residues. To obtain further information on how the C14 metacaspases I and II could have evolved from algae to angiosperms, their amino acid sequences were aligned by MUSCLE (see Supplementary data 2), and phylogenetic trees were constructed by the maximum likelihood PhyML method (Fig. 4). Two proteins truncated in the region between the catalytic His and Cys residues were excluded (RcMC7 and SbMC11) from this analysis. An alignment of the regions surrounding the putative location of the His and Cys catalytic residues for metacaspases I (Fig. 3b) showed that all proteins conserved the catalytic histidine residue with the exception of one rice and two sorghum proteins (OsMC1, SbMC3, and SbMC5), which change the histidine to leucine, isoleucine, or arginine, and the three metacaspases-I from *Ostreococcus*, in which the histidine is changed to arginine. Likewise, the conserved cysteine was changed to glutamic acid or glycine in OsMC1 and SbMC4 and to tyrosine or phenylalanine in the *Ostreococcus* proteins. The alignment of the amino acid regions surrounding the catalytic residues showed that both of them were conserved in all the metacaspase II proteins from algae to angiosperms (Fig. 3c). The phylogenetic tree constructed from metacaspase I proteins showed that the metacaspases from Prasinophyceae grouped in a separated clade. The C14 metacaspase I from Chlorophyceae and Trebouxiophyceae grouped in a clade that belongs to a cluster that contains putative orthologous proteins from algae, moss, spikemoss, and angiosperms, although species-specific duplication events make it difficult to establish actual orthologous relationships. A subset of this clade that contains only angiosperm proteins was also found. This second group was originated by a duplication event in the ancestor of angiosperm species, and it was expanded due to posterior duplication

events that are species-specific or in the ancestor of cereal sequences. Finally, one metacaspase from *Selaginella* with the most divergent amino acid sequence did not group with any other protein. The phylogenetic tree corresponding to metacaspases II showed that the C14 metacaspase II proteins from Chlorophyceae are ancestors to the other sequences. The rest of proteins could be distributed in three clades supported by aLRT values higher than 80%. One clade was formed by proteins of the moss, the spikemoss, and all the angiosperm species. A second clade was formed only by angiosperm sequences, which was presumably originated after a duplication event in the angiosperm ancestor. Finally, several moss metacaspases II did not group with proteins from other plant species.

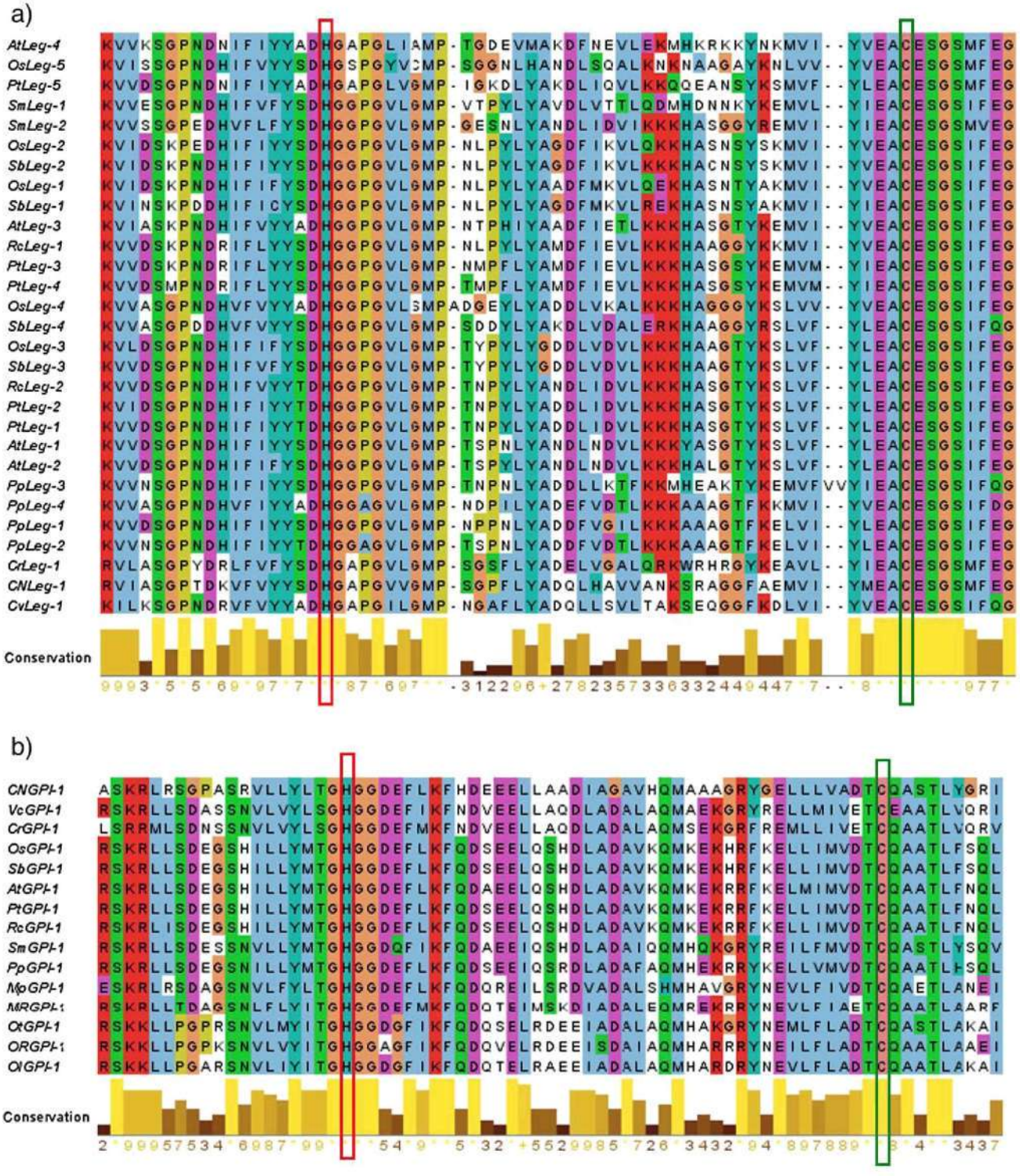


Fig. 1. Alignment of amino acid regions surrounding the catalytic His and Cys residues of C13 peptidases. (a) Legumains. (b) GPI:protein transamidases. Alignments were generated using the MUSCLE program. The putative locations of catalytic His (red box) and Cys (green box) residues are indicated. Ot, *O. tauri*, Ol, *O. lucimarinus*, OR, *Ostreococcus* sp. RCC809; Mp, *M. pusilla* CCMP1545; MR, *Micromonas sp.* RCC209; Cv, *C. vulgaris* C-169; CN, *Chlorella* sp. NC64A; Cr, *C. reinhardtii*; Vc, *V. carteri*; Pp, *P. patens*; Sm, *S. moellendorffii*; At, *A. thaliana*; Pt, *P. trichocarpa*; Rc, *R. communis*; Sb, *S. bicolor*; Os, *O. sativa*.

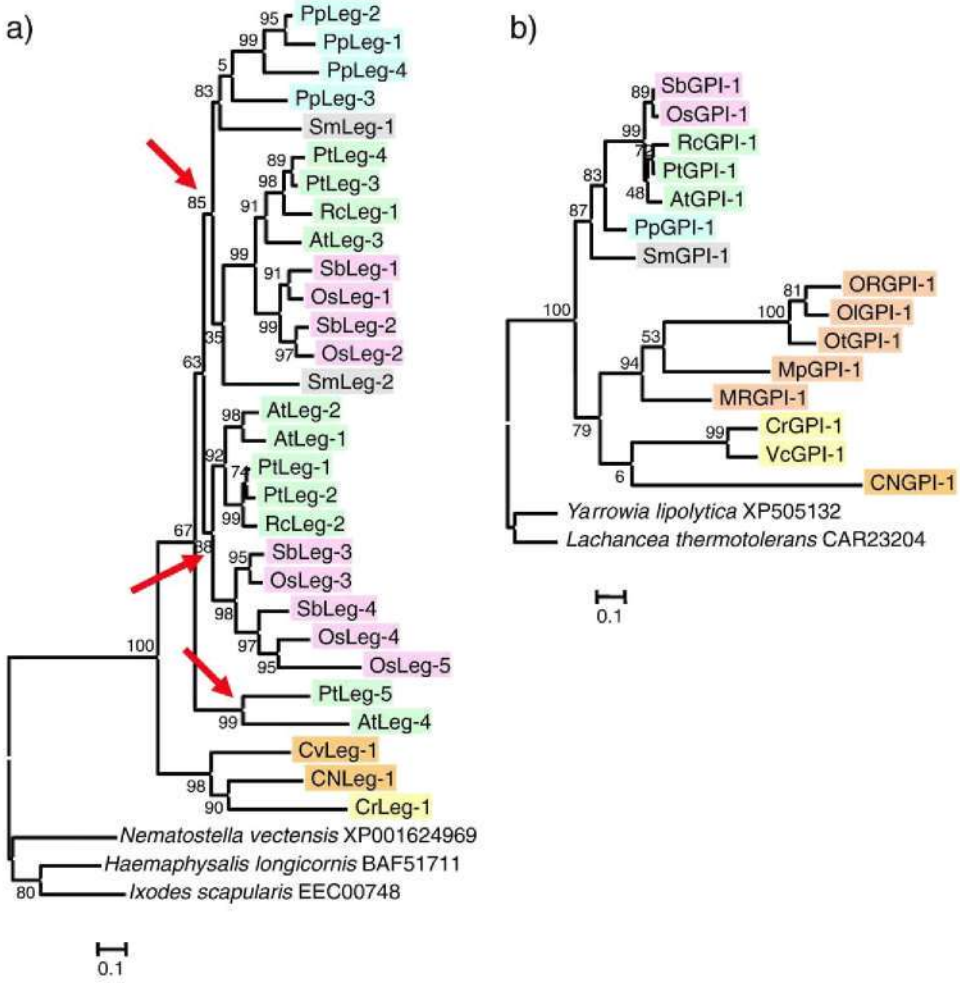


Fig. 2. Phylograms of the C13 peptidases from algae to angiosperms. (a) Legumains. (b) GPI:protein transamidases. The amino acid sequences were aligned by MUSCLE and analyzed with the PhyML method. Approximate likelihood ratio test values are indicated. Green, dicot proteins; pink, monocot proteins; blue, moss proteins; grey, spikemoss proteins; yellow, Chlorophyceae algal proteins; orange, Trebouxiophyceae algal proteins; tan, Prasinophyceae algal proteins. Red arrows mark nodes that are discussed in the text. Abbreviations as in Fig. 1.

3.4. Evolution of C50 separases in plants

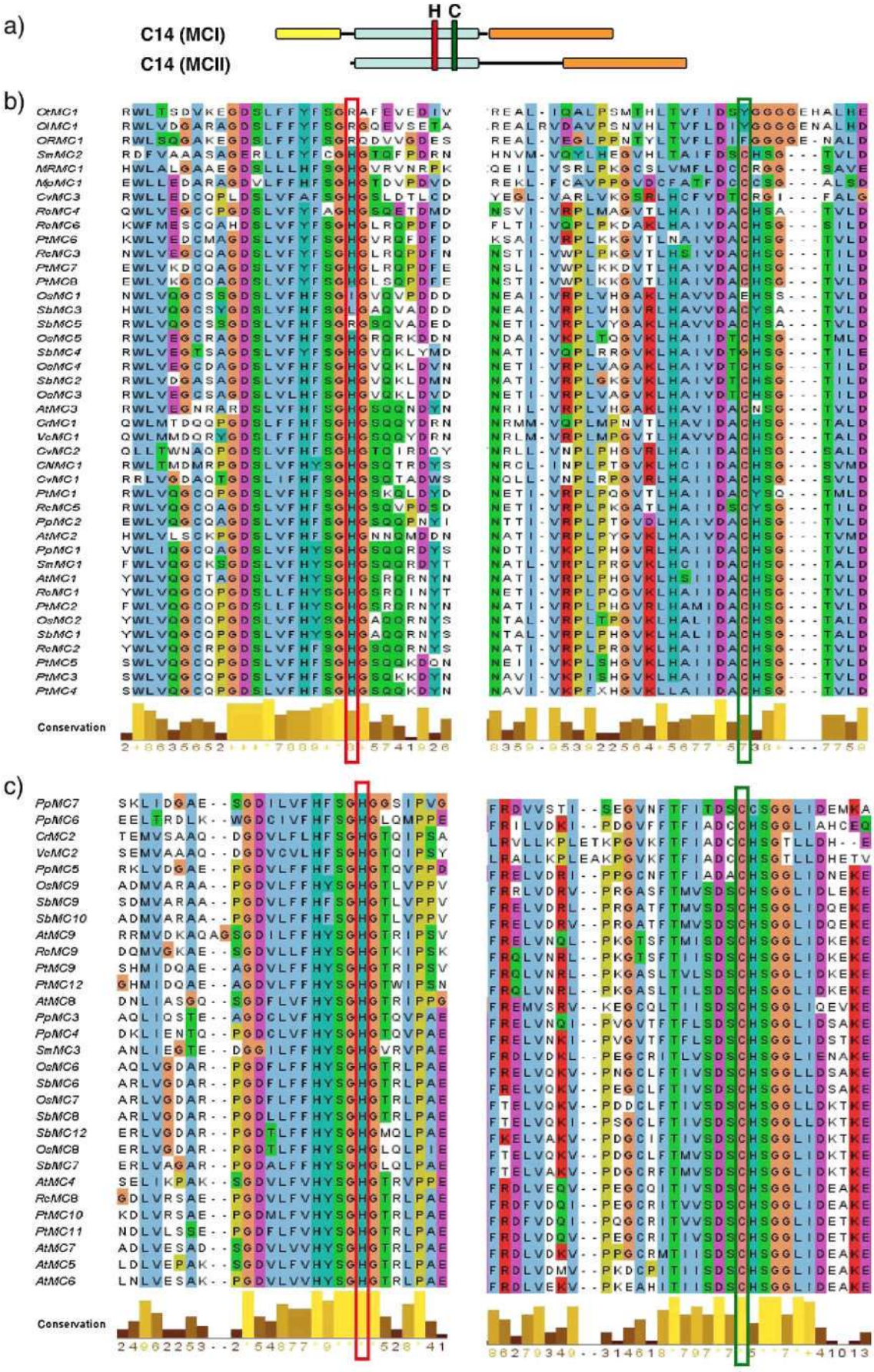
As performed to C13 and C14 proteases, the complete amino acid sequences from plant C50 separases were aligned by MUSCLE (see Supplementary data 2), and a phylogeny was constructed by the maximum likelihood PhyML method. The putative protein models from *O. tauri* and *Ostreococcus* sp. RCC809, truncated in the region between the catalytic His and Cys residues, were excluded from this analysis. The ORSep-1 protein corresponds to an EST cluster. Fig. 5a shows an alignment of the amino acid region that contains the putative His and Cys catalytic residues. Both amino acids are conserved in all sequences, which are highly similar in this region. The phylogenetic tree depicted in Fig. 5b almost resembles the evolutionary relationships among plant species, from algae to angiosperms. The separases from Prasinophyceae were the most divergent proteins. The Chlorophyceae and Trebouxiophyceae sequences were in the same group as proteins from land species but cluster in different clades. The moss and spikemoss separases grouped together. Finally, the angiosperm proteins are in the same cluster, but separases from monocot and dicot species group in different clades.

3.5. Evolution of C11 clostripains

After an extensive search in plant databases, only clostripain genes were found in the Chlorophyceae *V. carteri* and *C. reinhardtii*. To obtain further insights on the origin of these genes, the C11 clostripains from algae and different bacteria, archaea, and protists were aligned by MUSCLE (see Supplementary data 2), and a phylogeny was constructed by the maximum likelihood PhyML method. The alignment of the amino acid regions surrounding the catalytic His and Cys (typically separated for 40–45 amino acid residues) shows little amino acid conservation among clostripain sequences, but most of them present the specific clostripain motifs D/NHG and ACL (Fig. 6a). Phylogenetic analysis shows the algae clostripain sequences located in the same clade that the two eukaryotic proteins from the protists belonging to the Alveolata group *Cryptosporidium parvum* and *Toxoplasma gondii*, the archaea *Methanosarcina acetivorans*, and the bacteria *Syntrophobacter fumaroxidans*. The rest of bacterial sequences and a clostripain protein from the archaean *Methanospirillum hungatei* grouped in separated clades.

4. Discussion

The advent of extensive genome sequencing has lead comparative genomic analysis as a powerful tool to discover evolutionary relationships between protein families. Among the broad spectrum of peptidases, cysteine peptidases are crucial in different plant physiological processes. The clan CA of cysteine peptidases is the most known, and different works have described the evolutionary relationship of the papain-like C1A family (Beers et al., 2004; Garcia-


a)
H C
C14 (MCI)
C14 (MCII)
b)
OtMC1
OlMC1
ORMC1
SmMC2
MRMC1
MpMC1
CvMC3
RcMC4
RcMC6
PtMC6
RcMC3
PtMC7
PtMC8
OsMC1
SbMC3
SbMC5
OsMC5
SbMC4
OsMC4
SbMC2
OsMC3
AtMC3
GrMC1
VcMC1
CvMC2
CNMC1
CvMC1
PtMC1
RcMC5
PpMC2
AtMC2
PpMC1
SmMC1
AtMC1
RcMC1
PtMC2
OsMC2
SbMC1
RcMC2
PtMC5
PtMC3
PtMC4
Conservation
c)
PpMC7
PpMC6
CrMC2
VcMC2
PpMC5
OsMC9
SbMC9
SbMC10
AtMC9
RcMC9
PtMC9
PtMC12
AtMC8
PpMC3
PpMC4
SmMC3
OsMC6
SbMC6
OsMC7
SbMC8
SbMC12
OsMC8
SbMC7
AtMC4
RcMC8
PtMC10
PtMC11
AtMC7
AtMC5
AtMC6
Conservation

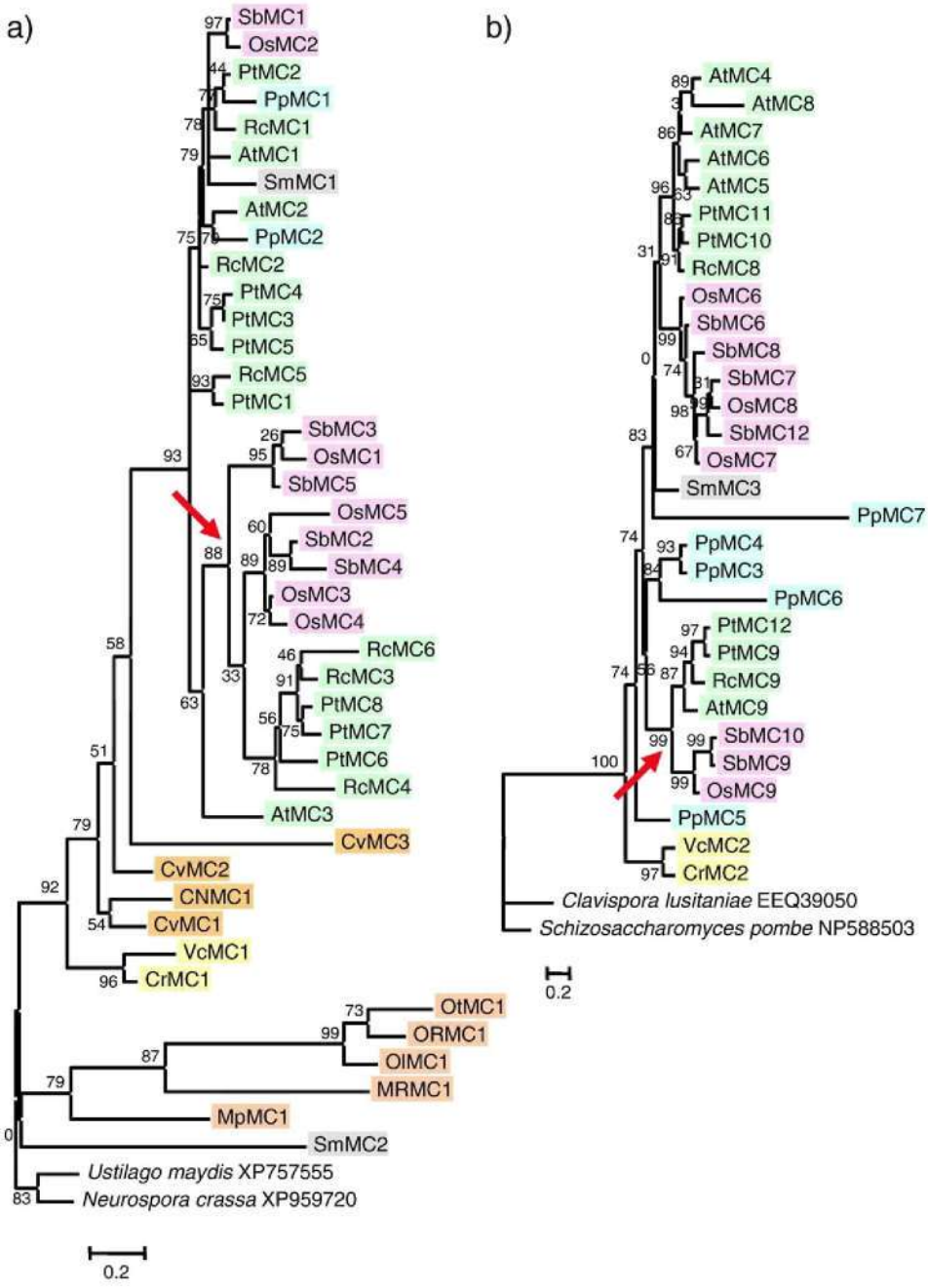


Fig. 4. Phylograms of the C14 peptidases from algae to angiosperms. (a) Type I metacaspases. (b) Type II metacaspases. The amino acid sequences were aligned by MUSCLE and analyzed with the PhyML method. Approximate likelihood ratio test values are indicated. Green, dicot proteins; pink, monocot proteins; blue, moss proteins; grey, spikemoss proteins; yellow, Chlorophyceae algal proteins; orange, Trebouxiophyceae algal proteins; tan, Prasinophyceae algal proteins. Red arrows mark nodes that are discussed in the text. Abbreviations as in Fig. 1.

Lorenzo et al., 2006; Martinez and Diaz, 2008). The clan CD has recently emerged as an important group of peptidases, but evolutionary relationships for this clan in plants have only been reported for the C13 legumain peptidases (Martinez and Diaz, 2008).

As expected, since the existence of plant members of these protein families had been previously described (Muntz and Shutov, 2002; Liu and Makaroff, 2006; Vercammen et al., 2007), C13 legumains and GPI:protein transamidases, C14 metacaspases I and II, and C50 separases were found when we made searches for clan CD peptidases in plants. Members of the bacterial families C25 gingipains and RTX toxins were not detected, but, unexpectedly, we discovered the existence of C11 clostripain members in two Chlorophyceae algae. Assigning the peptidases into the different protein families was easy, since they shared typical

Fig. 3. Sequence comparison of C14 metacaspases. (a) Schematic representation of the protein architectures of C14 metacaspases I and II. The catalytic domain is formed by the p20 (blue) and p10 (orange) subunits. The prodomain of type I metacaspases is in yellow. Positions of the catalytic His (H, red box) and Cys (C, green box) residues are indicated. (b) Alignment of the amino acid regions surrounding the catalytic His and Cys residues for type I metacaspases. (c) Alignment of the amino acid regions surrounding the catalytic His and Cys residues for type II metacaspases. Alignments were generated using the MUSCLE program. The putative locations of catalytic His (red boxes) and Cys (green boxes) residues are indicated. Abbreviations as in Fig. 1.

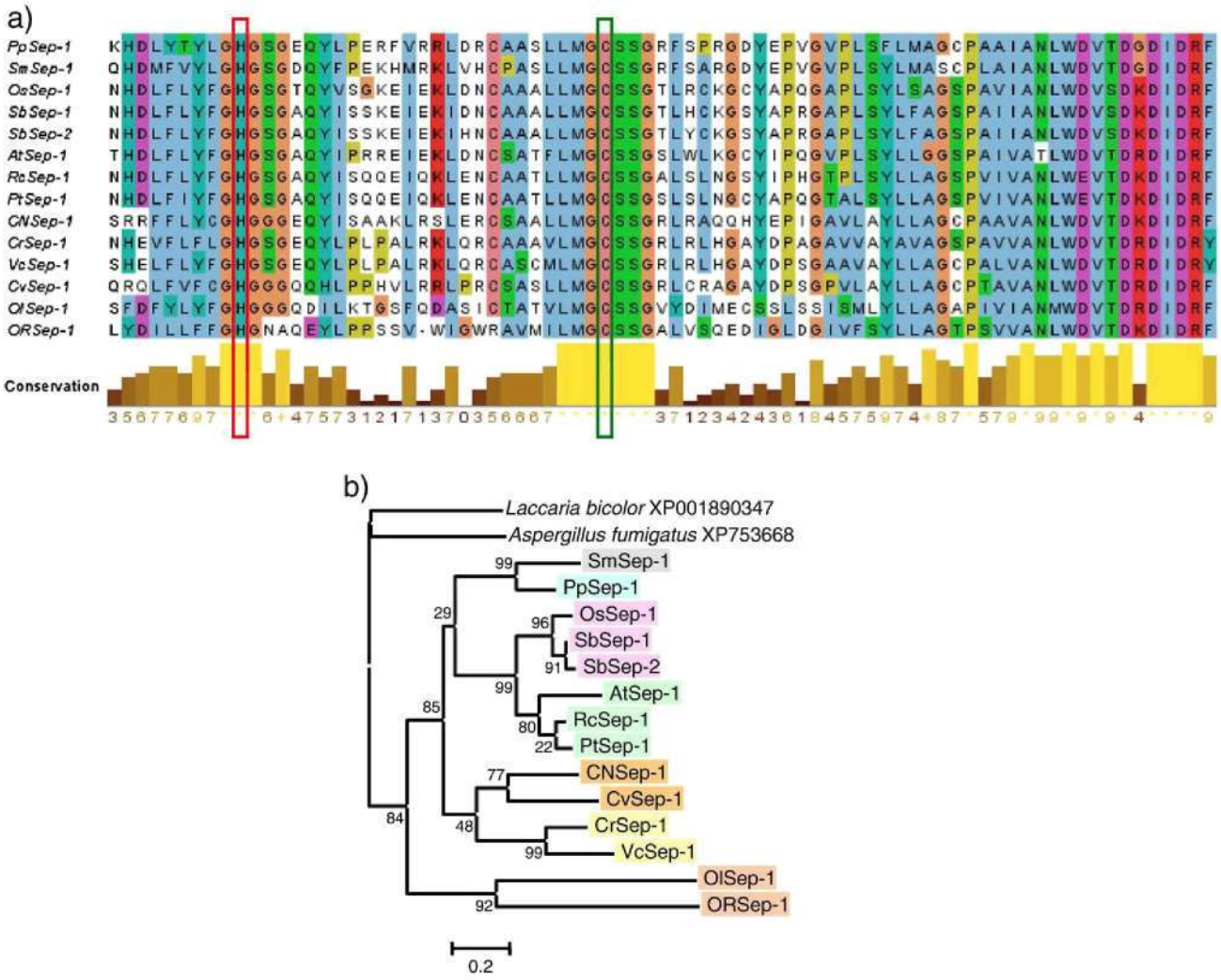


Fig. 5. Phylogenetic analysis of C50 separases. (a) Alignment of the amino acid regions surrounding the catalytic His and Cys residues (boxed) for C50 separases. Alignments were generated using the MUSCLE program. The putative locations of catalytic His (red box) and Cys (green box) residues are indicated. (b) Phylogram of the C50 separases from algae to angiosperms. The amino acid sequences were aligned by MUSCLE and analyzed with the PhyML method. Approximate likelihood ratio test values are indicated. Green, dicot proteins; pink, monocot proteins; blue, moss protein; grey, spikemoss protein; yellow, Chlorophyceae algal proteins; orange, Trebouxiophyceae algal proteins ; tan, Prasinophyceae algal proteins. Abbreviations as in Fig. 1.

protein architectures, and when a complete protein was not found, as it can be the result of in silico gene prediction by automatic methods, they present conserved signatures in the region surrounding the His and Cys catalytic residues. In this way, C13 legumains typically have the amino acid motifs DHG ACE, C13 GPI: protein transamidases GHG TCQ, C14 metacaspases I GHG ACH, C14 metacaspases II GHG SCH, C50 separases GHG GCS, and C11 clostripains D/NHG ACL.

The clostripain family is very diverse, with proteins that highly vary in length, protein architecture, and amino acid conservation (Labrou and Rigden, 2004). This variability probably reflects a range of physiological roles for these kinds of proteins. For the first time, we describe the existence of proteins belonging to this family of Viridiplantae. With respect to clostripain evolution, it seems that clostripains are ancestral genes present in primitive prokaryotes, since they have been found in archaea species and they have been maintained in some bacterial clades. Their presence in eukaryotes (green algae and apicomplexans) could be due to three mechanisms, namely, vertical gene transfer, lateral gene transfer, or endosymbiotic gene transfer. The fact that clostripain proteins have only been found in green algae and apicomplexan species leads us to think that these species have got these genes by endosymbiotic gene transfer. Multiple lines of evidence support the single origin of the primary plastid in the Plantae common ancestor, which was then transferred to chromalveolates (including apicomplexans) via secondary endosymbiosis (Reyes-Prieto et al., 2007). Moreover, endosymbiotic gene transfer is supported by the fact that, in most cases, algal endosymbiont nucleus is reduced (as in green algae), or has degenerated (as in apicomplexans), leaving behind a number of genes that have been transferred into the host nucleus (Archibald and Keeling, 2002; Reyes-Prieto et al., 2008). The reasons of their lack in most plant and chromalveolate lineages remain unknown.

The rest of the families of clan CD of peptidases might be divided in two groups. The first group would be formed by families present in all lineages, as is the case of C13 GPI:protein transamidases, C14 metacaspases I, and C50 separases. The second group would enclose C13 legumains and C14 metacaspases II, which are not present in the Prasinophyceae.

Peptidases from the first group must be key proteins in crucial processes in all the lineages tested. Post-translational modification with a GPI lipid anchor is an important mechanism for tethering proteins of eukaryotic organisms to the plasma membrane (Eisenhaber et al., 2003a; Zacks and Garg, 2006). The modification is executed by the transamidase complex located at the luminal side of the endoplasmatic

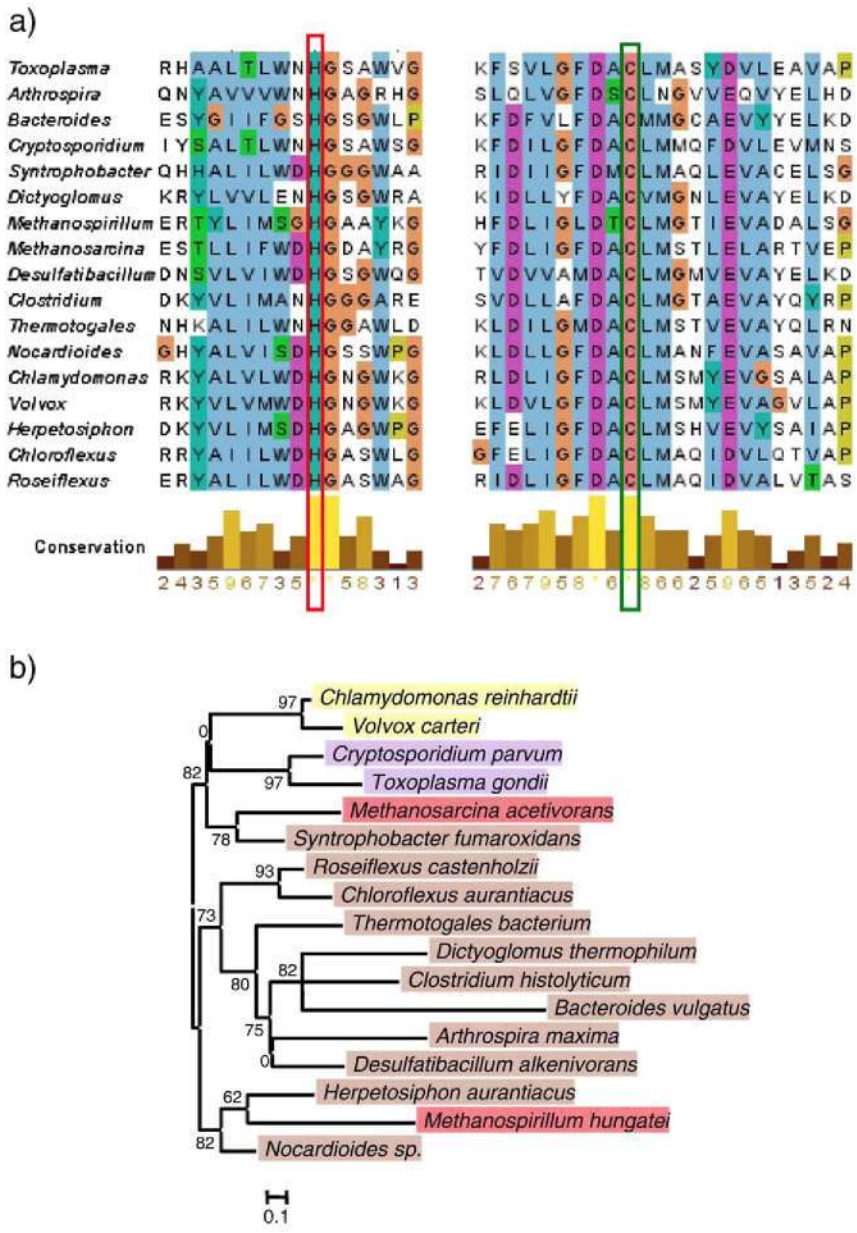


Fig. 6. Phylogenetic analysis of C11 clostripains. (a) Alignment of the amino acid regions surrounding the catalytic His and Cys residues (boxed) for C11 clostripains. Alignments were generated using the MUSCLE program. The putative locations of catalytic His (red box) and Cys (green box) residues are indicated. (b) Phylogram of the C14 separases from archaea, bacteria, protist, and algae. The amino acid sequences were aligned by MUSCLE and analyzed with the PhyML method. Approximate likelihood ratio test values are indicated. Pale red, archaeal proteins; brown, bacterial proteins; purple, protist proteins; yellow, algal proteins.

reticulum, in which the GPI:protein transamidases are the peptidases, and the processed protein is ultimately trafficked to the cell surface (Orlean and Menon, 2007). The modified proteins have an important role in cell–cell signalling in *Arabidopsis*, where they have been implicated in polarized cell expansion in the root (Schindelman et al., 2001), signal transduction during disease resistance (Coppinger et al., 2004), and male and female gametogenesis (Acosta-Garcia and Vielle-Calzada, 2004; Capron et al., 2008). Then, it is expected that members of this peptidase family exist in all plant species. This gene was found in all organisms tested with the exception of one of the *Chlorella* strains. The fact that the other *Chlorella* strain possess one GPI:protein transamidase gene leads us to think that imperfect genome sequencing could explain why this gene was not observed. Evolutionarily, one gene is present in each plant species, supporting its ancestral role.

The case of separases is very similar. Separases are mainly involved in the release of chromosome cohesion at the metaphase–anaphase transition and the subsequent separation of sister chromatids by cleaving of the cohesin kleisin (SCC1) subunit during eukaryotic mitosis and meiosis (Queralt and Uhlmann, 2005; Liu and Makaroff, 2006). Separases have been implicated in some other processes as in anaphase spindle stabilization or in cell cycle progression (Papi et al., 2005; Queralt and Uhlmann, 2005, 2008). As cell cycle is a general process, separases are expected to be present in all plant species. Surprisingly, in the algae *Micromonas*, we have not found genes encoding separase proteins, which also occur in some *Plasmodium* and *Theileria* species as deduced from genomic database searches. Two possibilities arise: separases are not present in these organisms and their role is achieved by other proteins, or these genes were not observed due to imperfect sequencing. Plant evolution has not involved the duplication of these genes, since only one is maintained in all land plants, with the exception of sorghum. The two proteins of sorghum have more than 85% identity, in

contrast to a percentage of identity about 70% of both proteins with its putative orthologous protein from rice. This supports that a recent duplication event has lead to the appearance of a second separase gene in sorghum.

Metacaspases I are also maintained in all lineages of algae to angiosperms, but unlike separases or GPI:protein transamidases, the number of homologs increases in land plants (intriguingly, there are three metacaspases I in one *Chlorella* strain). Metacaspase I sequences have been reported among protozoans, fungi, algae, and plants (Vercammen et al., 2007), which are derived from the metacaspase-like sequences reported in bacteria (Bidle and Falkowski, 2004). Metacaspases II have been only reported in plants and in the choanoflagellate *Monosiga brevicollis* (which probably acquired it by lateral gene transfer from a photosynthetic alga). Strikingly, type II metacaspases are not present in the Prasinophyceae and Trebouxiophyceae algae. Although the numbers of metacaspases I and II slightly increase in land plants, there is no correlation between the numbers of each type of metacaspases in each lineage. For example, there are more metacaspases I (8) than metacaspases II (4) in poplar, whereas the reverse is observed in *Arabidopsis* (3 and 6, respectively). It can be deduced from the phylogenetic trees that duplications following speciation, leading to neofunctionalization, subfunctionalization, or redundancy of these proteins, together with duplication events occurring in the ancestor of angiosperm species, are the responsible forces for this variability. Until now, the role of plant metacaspases is poorly documented. Caspases are responsible for apoptosis in animals (Chowdhury et al., 2008). In plants, caspase-like activities have been reported to be involved in programmed cell death (Bonneau et al., 2008). Plant genomes do not contain structural homologs of caspases and metacaspases have been proposed to be involved in the regulation of these programmed cell death processes (Uren et al., 2000). In tomato, the expression of a type II metacaspase increases upon infection with the fungus *Botrytis cinerea*, (Hoeberichts et al., 2003) and knocking out type II metacaspases in Norway spruce abolishes somatic embryogenesis-related cell death (Suarez et al., 2004; Bozhkov et al., 2005). Likewise, an in silico analysis of publicly available microarray resources suggests that some metacaspases might be involved in cell death-related processes (Sanmartin et al., 2005). However, despite conservation of the catalytic dyad of histidine and cysteine in metacaspases and caspases, their overall sequence similarity is very low and their substrate specificity differs because caspases have an Asp specific proteolytic activity and metacaspases of plants, fungi, and protozoa have shown an Arg/Lys-specific proteolytic activity (Vercammen et al., 2004, 2006; Bozhkov et al., 2005; Watanabe and Lam, 2005; Gonzalez et al., 2007). Thus, metacaspases could not be directly involved in the regulation of cell death but rather, directly or indirectly, in signalling cascades leading to cell death (Vercammen et al., 2007). Currently, there is no evidence of whether the two types of plant metacaspases act in the same pathways in plant cells. Evolutionary analyses suggest an ancestral role for metacaspases that group with algae sequences in the phylogenetic tree, and new roles could have been acquired by type I and type II metacaspases located in separated clades formed only by angiosperm sequences. The functionality of Prasinophyceae sequences is controversial, since *Ostreococcus* proteins lack the catalytic residues whereas they are maintained in *Micromonas* proteins.

Finally, as reported in Martinez and Diaz (2008), legumain-like cysteine peptidases are not present in Prasinophyceae algae. In higher plants, legumains are involved in important physiological processes. They are implicated in the activation of papain-like cysteine peptidases to degrade storage proteins in the seed (Okamoto and Minamikawa, 1999; Kato et al., 2003) and in defense against pathogens executing programmed cell death (Hatsugai et al., 2004; Rojo et al., 2004). An important but restricted function of these peptidases could account for the small number of members of this protein family. The duplication events leading to a higher number of legumain proteins in angiosperms suggest that new physiological roles could have been acquired in these species.

Overall, the main differences in clan CD peptidases are restricted to algae clades. The green plants are divided into the phyla Streptophyta and Chlorophyta. The Streptophyta contains all land plants and the green algae belonging to the class Charophyceae, from which, unfortunately, no genome has been completely sequenced. The Chlorophyta contains the members of the classes Prasinophyceae, Ulvophyceae, Trebouxiophyceae, and Chlorophyceae (Lewis and McCourt, 2004), of which the Prasinophyceae are the most basal. Comparison of the algal genome sequences revealed that individual organisms possess unique features in terms of primary sequence structures and gene compositions, making each useful for understanding basic physiological aspects and the evolution of photosynthetic eukaryotes (Misumi et al., 2008). Comparative analysis of the genomes of two *Ostreococcus* species has revealed major differences in genome organization between them, which may reflect ongoing adaptation and speciation processes (Palenik et al., 2007). In addition, both *Ostreococcus* species employ similar mechanisms for optimization of genome and cell size, including gene loss, gene fusion, utilization of selenocysteine-containing proteins, and chromatin reduction. The number of gene models could be a clue to deal with the great differences in clan CD families in algae. The genomes in Prasinophyceae and Trebouxiophyceae are smaller than in Chlorophyceae, which could be the result of an adaptive process to the environment and could account for the loss of clan CD genes that did not perform a crucial role in these organisms.

In conclusion, comparative genomic analyses have provided us valuable insights into the conservation and evolution of the different families included in the clan CD of cysteine peptidases in plants. A phylogenetic analysis of these gene families in representative species of different plant taxonomic groups has permitted us to state that whereas C13 GPI:protein transamidases, C14 metacaspases I, and C50 separases are present in all groups, C13 legumains and C14 metacaspases II are absent in some basal groups, probably due to adaptation to environment. Moreover, C11 clostripains have been found in some species of Chlorophyceae algae. The low number of protein members in these families suggests a limited and conserved physiological role for these proteins. Late gene duplication events mainly observed in land plants could be associated with the increasing structural and functional complexities acquired.

Given that all the sequences discussed share a common fold, it must be possible to come up with a general evolutionary scheme explaining in which organism kingdoms particular families and subfamilies arose. Discounting the odd sequences that may be derived by lateral gene transfer, C13 and C14B would appear to be the oldest (sub)families, perhaps present in the ultimate ancestor, C11 present in the ancestor of bacteria and archaea, C14A and C50 present in the ancestor of the eukaryotes, and C25 and C80 present only in the ancestor of bacteria. The absence of C14A homologs in plants could be explained by gene loss.

Acknowledgments

The financial support from the Ministerio de Educación y Ciencia (BFU2008-01166) is gratefully acknowledged. Thanks to Dr I. Diaz for critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2009.09.003.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Acosta-Garcia, G., Vielle-Calzada, J.P., 2004. A classical arabinogalactan protein is essential for the initiation of female gametogenesis in *Arabidopsis*. Plant Cell 16, 2614–2628.

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552.

Aravind, L., Koonin, E.V., 2002. Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. Proteins 46, 355–367.

Archibald, J.M., Keeling, P.J., 2002. Recycled plastids: a 'green movement' in eukaryotic evolution. Trends Genet. 18, 577–584.

Baum, P., Yip, C., Goetsch, L., Byers, B., 1988. A yeast gene essential for regulation of spindle pole duplication. Mol. Cell Biol. 8, 5386–5397.

Beers, E.P., Jones, A.M., Dickerman, A.W., 2004. The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. Phytochemistry 65, 43–58.

Belenghi, B., et al., 2007. Metacaspase activity of *Arabidopsis thaliana* is regulated by S-nitrosylation of a critical cysteine residue. J. Biol. Chem. 282, 1352–1358.

Bidle, K.D., Falkowski, P.G., 2004. Cell death in planktonic, photosynthetic microorganisms. Nat. Rev. Microbiol. 2, 643–655.

Bonneau, L., Ge, Y., Drury, G.E., Gallois, P., 2008. What happened to plant caspases? J. Exp. Bot. 59, 491–499.

Bozhkov, P.V., et al., 2005. Cysteine protease mcII-Pa executes programmed cell death during plant embryogenesis. Proc. Natl. Acad. Sci. U. S. A. 102, 14463–14468.

Capron, A., et al., 2008. Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the LORELEI gene. Plant Cell 20, 3038–3049.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Coppinger, P., Repetti, P.P., Day, B., Dahlbeck, D., Mehlert, A., Staskawicz, B.J., 2004. Overexpression of the plasma membrane-localized NDR1 protein results in enhanced bacterial disease resistance in *Arabidopsis thaliana*. Plant J. 40, 225–237.

Chen, J.M., Rawlings, N.D., Stevens, R.A., Barrett, A.J., 1998. Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases. FEBS Lett. 441, 361–365.

Chowdhury, I., Tharakan, B., Bhat, G.K., 2008. Caspases—an update. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 151, 10–27.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Eichinger, A., et al., 1999. Crystal structure of gingipain R: an Arg-specific bacterial cysteine proteinase with a caspase-like fold. EMBO J. 18, 5453–5462.

Eisenhaber, B., Maurer-Stroh, S., Novatchkova, M., Schneider, G., Eisenhaber, F., 2003a. Enzymes and auxiliary factors for GPI lipid anchor biosynthesis and post-translational transfer to proteins. Bioessays 25, 367–385.

Eisenhaber, B., Wildpaner, M., Schultz, C.J., Borner, G.H., Dupree, P., Eisenhaber, F., 2003b. Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for *Arabidopsis* and rice. Plant Physiol. 133, 1691–1701.

Garcia-Lorenzo, M., Sjodin, A., Jansson, S., Funk, C., 2006. Protease gene families in *Populus* and *Arabidopsis*. BMC Plant Biol. 6, 30.

Gonzalez, I.J., Desponds, C., Schaff, C., Mottram, J.C., Fasel, N., 2007. *Leishmania major* metacaspase can replace yeast metacaspase in programmed cell death and has arginine-specific cysteine peptidase activity. Int. J. Parasitol. 37, 161–172.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hara-Nishimura, I., Inoue, K., Nishimura, M., 1991. A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. FEBS Lett. 294, 89–93.

Hatsugai, N., et al., 2004. Plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. Science 305, 855–858.

Hatsugai, N., Kuroyanagi, M., Nishimura, M., Hara-Nishimura, I., 2006. A cellular suicide strategy of plants: vacuole-mediated cell death. Apoptosis 11, 905–911.

Hoeberichts, F.A., ten Have, A., Woltering, E.J., 2003. A tomato metacaspase gene is upregulated during programmed cell death in *Botrytis cinerea*-infected leaves. Planta 217, 517–522.

Kato, H., Sutoh, K., Minamikawa, T., 2003. Identification, cDNA cloning and possible roles of seed-specific rice asparaginyl endopeptidase, REP-2. Planta 217, 676–685.

Labrou, N.E., Rigden, D.J., 2004. The structure–function relationship in the clostripain family of peptidases. Eur. J. Biochem. 271, 983–992.

Lewis, L., McCourt, R., 2004. Green algae and the origin of land plants. Am. J. Bot. 91, 1535–1556.

Liu, Z., Makaroff, C.A., 2006. *Arabidopsis* separase AESP is essential for embryo development and the release of cohesin during meiosis. Plant Cell 18, 1213–1225.

Lupardus, P.J., Shen, A., Bogyo, M., Garcia, K.C., 2008. Small molecule-induced allosteric activation of the *Vibrio cholerae* RTX cysteine protease domain. Science 322, 265–268.

Martinez, M., Diaz, I., 2008. The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. BMC Evol. Biol. 8, 198.

Misumi, O., et al., 2008. Genome analysis and its significance in four unicellular algae, *Cyanidioshyzon merolae*, *Ostreococcus tauri*, *Chlamydomonas reinhardtii*, and *Thalassiosira pseudonana*. J. Plant Res. 121, 3–17.

Muntz, K., Shutov, A.D., 2002. Legumains and their functions in plants. Trends Plant Sci. 7, 340–344.

Nedelcu, A.M., Miles, I.H., Fagir, A.M., Karol, K., 2008. Adaptive eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. J. Evol. Biol. 21, 1852–1860.

Okamoto, T., Minamikawa, T., 1999. Molecular cloning and characterization of *Vigna mungo* processing enzyme 1 (VmPE-1), an asparaginyl endopeptidase possibly involved in post-translational processing of a vacuolar cysteine endopeptidase (SH-EP). Plant Mol. Biol. 39, 63–73.

Orlean, P., Menon, A.K., 2007. Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. J. Lipid Res. 48, 993–1011.

Palenik, B., et al., 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. U. S. A. 104, 7705–7710.

Papi, M., Berdougo, E., Randall, C.L., Ganguly, S., Jallepalli, P.V., 2005. Multiple roles for separase auto-cleavage during the G_2/M transition. Nat. Cell Biol. 7, 1029–1035.

Queralt, E., Uhlmann, F., 2005. More than a separase. Nat. Cell Biol. 7, 930–932.

Queralt, E., Uhlmann, F., 2008. Separase cooperates with Zds1 and Zds2 to activate Cdc14 phosphatase in early anaphase. J. Cell Biol. 182, 873–883.

Rawlings, N.D., Morton, F.R., Kok, C.Y., Kong, J., Barrett, A.J., 2008. MEROPS: the peptidase database. Nucleic Acids Res. 36, D320–D325.

Reyes-Prieto, A., Weber, A.P., Bhattacharya, D., 2007. The origin and establishment of the plastid in algae and plants. Annu. Rev. Genet. 41, 147–168.

Reyes-Prieto, A., Moustafa, A., Bhattacharya, D., 2008. Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. Curr. Biol. 18, 956–962.

Rojo, E., et al., 2004. VPEgamma exhibits a caspase-like activity that contributes to defense against pathogens. Curr. Biol. 14, 1897–1906.

Sanmartin, M., Jaroszewski, L., Raikhel, N.V., Rojo, E., 2005. Caspases. Regulating death since the origin of life. Plant Physiol. 137, 841–847.

Schindelman, G., et al., 2001. COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. Genes Dev. 15, 1115–1127.

Shimada, T., et al., 2003. Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*. J. Biol. Chem. 278, 32292–32299.

Suarez, M.F., et al., 2004. Metacaspase-dependent programmed cell death is essential for plant embryogenesis. Curr. Biol. 14, R339–R340.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Uhlmann, F., Wernic, D., Poupart, M.A., Koonin, E.V., Nasmyth, K., 2000. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell 103, 375–386.

Uren, A.G., et al., 2000. Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. Mol. Cell 6, 961–967.

Vercammen, D., et al., 2004. Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine. J. Biol. Chem. 279, 45329–45336.

Vercammen, D., et al., 2006. Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. J. Mol. Biol. 364, 625–636.

Vercammen, D., Declercq, W., Vandenabeele, P., Van Breusegem, F., 2007. Are metacaspases caspases? J. Cell Biol. 179, 375–380.

Walker, N.P., et al., 1994. Crystal structure of the cysteine protease interleukin-1 beta-converting enzyme: a (p20/p10)2 homodimer. Cell 78, 343–352.

Watanabe, N., Lam, E., 2005. Two *Arabidopsis* metacaspases AtMCP1b and AtMCP2b are arginine/lysine-specific cysteine proteases and activate apoptosis-like cell death in yeast. J. Biol. Chem. 280, 14691–14699.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191.

Zacks, M.A., Garg, N., 2006. Recent developments in the molecular, biochemical and functional characterization of GPI8 and the GPI-anchoring mechanism [review]. Mol. Membr. Biol. 23, 209–225.